

09/266680

ABSTRACT OF THE DISCLOSURE

A method of generating a definition for a collection of source documents is provided. Patterns common to each source document in the collection of source documents are identified and restrictive general rules based on the identified common patterns are then constructed for element types. The construction of a restricted general rule includes constructing a content model that specifies the sequence order and number of occurrences of sub-elements within the common pattern. It further includes constructing attribute definitions and values rules for attributes occurring in the common patterns. Also provided is a method of converting a format of a first source document to a format of a similarly structured second source document is provided. The method identifies patterns common to the first and second source documents and maps elements and sub-elements in common pattern of the first source document to equivalent elements and sub-elements in the common pattern of the second source document.